

Modeling and optimization of a continuous electrocoagulation process using an artificial intelligence approach

Nuno S. Graça ^{*}, Ana M. Ribeiro and Alírio E. Rodrigues

Laboratory of Separation and Reaction Engineering – Laboratory of Catalysis and Materials (LSRE-LCM), Department of Chemical Engineering, University of Porto, Rua Dr. Roberto Frias, s/n, 4200-465 Porto, Portugal

*Corresponding author. E-mail: nunopsg@fe.up.pt

 NSG, 0000-0002-4397-6578

ABSTRACT

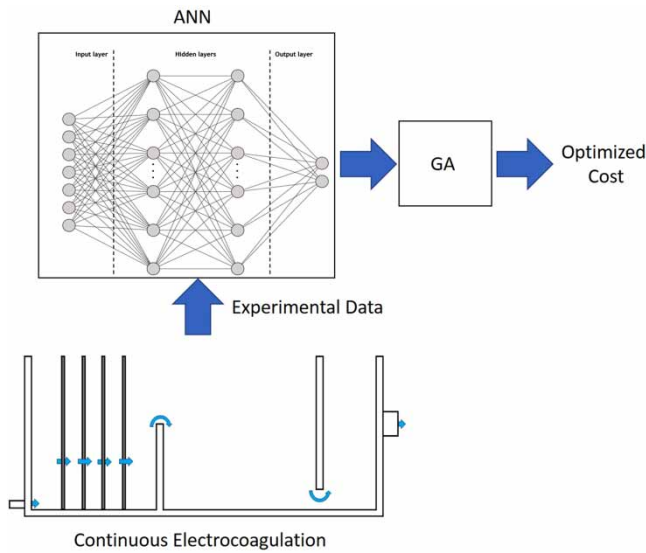
An artificial neural network (ANN) with the topology 8-94-85-2 (input – hidden layer 1 - hidden layer 2 - output) was used to model the operation of the continuous electrocoagulation (CEC) process for the removal of fluoride from water. After the ANN training, the sum of the squared errors (MSE) and the determination coefficient (R^2) of the testing set model predictions were 0.0088 and 0.999, respectively, showing a good generalization and the model's predictive capacity. The optimization of the process cost using the genetic algorithm (GA) showed that the optimal conditions are highly dependent on the feed concentration and the fluoride removal requirements. For 5 L of water containing 10 mg/L of fluoride, the optimal conditions to reduce the fluoride concentration below the permissible limit (1.5 mg/L) are 88.3 mA of current intensity, a flow rate of 73.6 mL/min, and the use of a series monopolar (SM) electrode configuration, corresponding to a fluoride removal of 85% and an operating cost of 0.05 €/L.

Key words: electrocoagulation, genetic algorithm, neural networks, water treatment

HIGHLIGHTS

- Using Open-Source Software for modeling the electrocoagulation process by a machine learning approach.
- Implantation of the genetic algorithm with an artificial neural network model to optimize the process.
- The impact of the optimized operating conditions on the process cost.
- The possibility of using hybrid processes to reduce the water treatment cost.

GRAPHICAL ABSTRACT



1. INTRODUCTION

Fluoride is considered a potentially dangerous species for human health, and the maximum permissible concentration of fluoride in drinking water has been set at 1.5 mg/l (World Health Organization 1993; Drinking Water Directive 1998). The continuous exposure to high fluoride levels in drinking water affects the metabolism of calcium, potassium, and phosphorus in the human body leading to several health problems (Fakhri & Adami 2013). Health issues such as dental, skeletal, and non-skeletal forms of fluorosis affect over 60 million people in India (Deepthi *et al.* 2021), and about 120 million people are in high-risk areas for fluoride exposure (Podgorski *et al.* 2018). The presence of high fluoride levels in the water resources is a worldwide problem affecting places such as China, Italy, the Middle East, Mexico, Netherlands, Norway, Pakistan, Poland, Spain, UK, and various regions of Africa (Mollah *et al.* 2004; Lacson *et al.* 2021). The development of efficient water defluoridation processes is then fundamental to reduce the impact of fluoride exposure on millions of people worldwide. In recent years, electrochemical techniques such as electrocoagulation (EC) have been considered an efficient option for water defluoridation (Changmai *et al.* 2018; Castañeda *et al.* 2020; Khan *et al.* 2020). The performance of the electrocoagulation process is affected by the nature of the pollutant and its concentration, the current intensity applied, electrodes material and its configuration (Fajardo *et al.* 2017; Silva *et al.* 2018). Therefore, the knowledge of the effect of the operating conditions on the EC performance is fundamental to optimize the efficiency of the process.

The mathematical modeling of a process allows a better grasp of how the different parameters and operating conditions affect its performance. This approach can help reduce the cost and the number of experiments required to optimize a process. However, in some processes, such as electrocoagulation, mathematical modeling is not a trivial task. The difficulty of modeling the electrocoagulation-based processes results from the complexity of the phenomena involved, such as electrode material speciation, polymerization reactions, different solubility of the generated coagulant species, different removal mechanisms, and electrodes passivation. Some mathematical models have been proposed for different polluted water systems (Matteson *et al.* 1995; Khemis *et al.* 2006; Lacasa *et al.* 2013; Graça *et al.* 2019a). However, a phenomenological model that satisfactorily describes the different processes involved in this kind of treatment is still hard to obtain (Cañizares *et al.* 2008).

Several methodologies and tools using artificial intelligence (AI) can be an alternative to replace the conventional phenomenological models for data-driven models such as artificial neural networks (ANN) (Nasr *et al.* 2016; Bock *et al.* 2019; Morales-Rivera *et al.* 2020). ANNs are advanced machine learning algorithms suitable for fitting and pattern recognition, allowing the extraction of complex relationships from a set of linear and nonlinear input variables to predict target outputs (Sivanandam & Deepa 2006). Machine learning algorithms have found application in a wide range of fields such as self-driving cars (Duarte & Ratti 2018), face recognition (Voulodimos *et al.* 2018), medical diagnosis (Borgli *et al.* 2019), and automated translation (Wu *et al.* 2016). The ANNs development and application in recent years have been taking advantage

of the faster, cheaper, and more powerful parallel processing capacity provided by the advent of graphic processing units (GPUs) (Jones *et al.* 2018). An ANN is constituted by a series of interconnected nodes (neurons) and, to some extent, tries to emulate the natural processing ability of the human brain (Barron 1993). The basic unit of the ANN is the artificial neuron, which processes information from one or more inputs, producing an output. A training algorithm is used in order to adjust the ANN parameters (weights and bias). One of the most used training algorithms is backpropagation with gradient descent optimization. In its most straightforward implementation, the backpropagation updates the weights and bias in the proportion and direction of the performance function gradient, minimizing the errors between the output values predicted by the network and the target output values (Wang *et al.* 2018). One of the main advantages of the ANNs is that their application is independent of a detailed knowledge of the process to be modeled. Moreover, the adjustment of the model parameters is simpler than in other conventional methods, and the prediction performance can be improved by updating the model with new training data (Prakasham *et al.* 2011).

Another AI-based tool is the genetic algorithm (GA). GA is a meta-heuristic optimization method based on the natural selection of the individuals (solutions), where the most adapted (best solutions) will have the best chance to reproduce, generating descendants better adapted than their parents (better solutions). By repeating the selection process and cross-breeding for several generations, it will be possible to find the best-adapted individual (the best solution). This individual represents the solution for a global optimum of a given objective function (Whitley 1994; Picos-Benítez *et al.* 2017). One of the GA's key features is its ability to exploit and gather information about an initially unknown search space, only requiring the definition of the upper and lower boundaries of the variables that define the search space. GA begins to search globally throughout the entire search space. The search is then directed to the region with a higher probability of providing a better solution through selection, crossover, and mutation (Lin 2004).

The implementation of machine learning algorithms to solve engineering problems is frequently carried out using commercial software such as Matlab® (Nogueira *et al.* 2018; Oliveira *et al.* 2020). However, alternative open-source software has been widely used in different scientific and technological fields. Open-source software is readily available to the users compared to commercial software, which can be highly expensive. Moreover, open-source tools have demonstrated their scalability through their frequent use in industry and academia (Inguva *et al.* 2021). Python is the most popular programming language in machine learning applications. Several open-source machine-learning libraries are based on Python, such as Tensorflow, PyTorch, Caffe, and Theano (Shen & Liu 2020). Tensorflow is an open-source machine learning framework initially created by researchers at Google, and it is one of the most widely used. Although there is widespread use of these open-source tools in fields such as informatics and electronics, their implementation in the chemical engineering field is still significantly reduced.

The present work aims to develop an ANN model able to predict the output concentration and the applied voltage of a CEC unit for the removal of fluoride from water. The ANN model implementation, training, and testing were carried out on the open-source Python Keras library with Tensorflow as a backend (Chollet 2015). The trained ANN model was used to construct the fitness function of a GA developed in Python, used to optimize the performance of the unit in terms of fluoride removal and operating costs.

2. MATERIALS AND METHODS

The implementation and training of an ANN to model a process such as CEC require a set of data containing the input variables and their respective outputs. The data obtained from the experimental work with the CEC unit for the removal of fluoride from water (Graça *et al.* 2019b) was used in the present work, where the operating parameters current intensity, flowrate, feed concentration, operation time, and electrodes configuration were used as input variables, and unit outlet concentration and the applied voltage were used as the outputs.

2.1. Continuous electrocoagulation reactor (CEC)

The continuous electrocoagulation unit is constituted by a plexiglass structure, $13 \times 26.6 \times 12.5$ cm, consisting of a first compartment containing the electrodes and directly connected to the unit inlet; a second compartment, which receives the water from the first compartment, is connected to a third compartment by a 1.5×12.5 cm gap in the bottom of the reactor. This separation avoids the floating solids from the second compartment passing directly to the reactor outlet. The four aluminum electrodes used in the experiments, $15 \times 10 \times 0.2$ cm, were connected to a DC power supply operating under galvanostatic conditions. The electrodes' arrangement inside the reactor makes the water flow in a serpentine pattern through the different

polarities of the electrodes (Figure 1). A more detailed description of the unit and the experimental procedures is available in Graça *et al.* (2019b).

The electrical connections between the electrodes and the power source can be arranged in three different configurations (Figure 2). Two of those configurations are monopolar, meaning that each electrode is charged with only one polarity. This kind of configuration can be series monopolar (SM), where the pair of internal electrodes is connected to each other but has no connection with the outer electrodes, or parallel configuration (P), where all the electrodes are connected in pairs to the power source. Alternatively, the electrodes can be arranged in a series bipolar configuration (SB), where only the outer electrodes are connected to the power source, and the internal electrodes are placed without any connection. In this configuration, the internal electrodes present different polarities in each face (Figure 2(b)).

2.2. Data pre-processing

The data used in the present work was taken from a previous work where a set of experiments designed using the Box-Behnken method was used to perform a statistical analysis of the process (Graça *et al.* 2019b). The design of the experiments considered three continuous factors (applied current, flow rate, and feed concentration) and one categorical factor (electrodes configuration). The application of the Box-Behnken design considering three repetitions of the central point resulting in 45 experiments. During the time (six sampling points), the outlet concentration of the CEC unit and the applied voltage necessary to maintain the current intensity were measured for each experiment. From this data, it was determined that the final

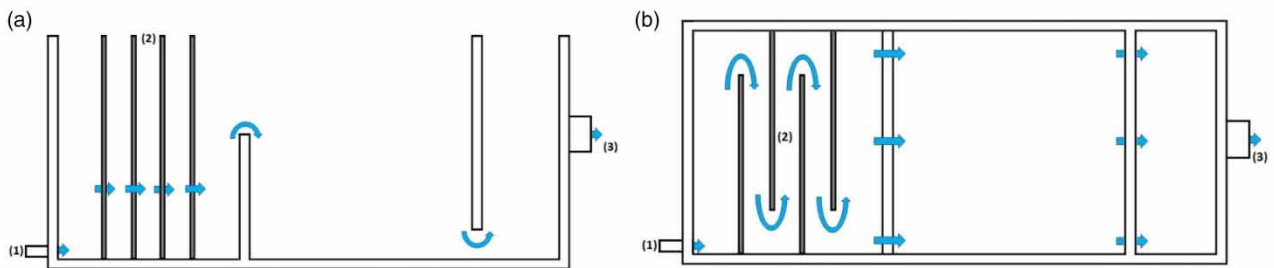


Figure 1 | Schematic representation of the continuous electrocoagulation reactor: (a) side view; (b) top view (1- inlet, 2- electrodes, 3- outlet).

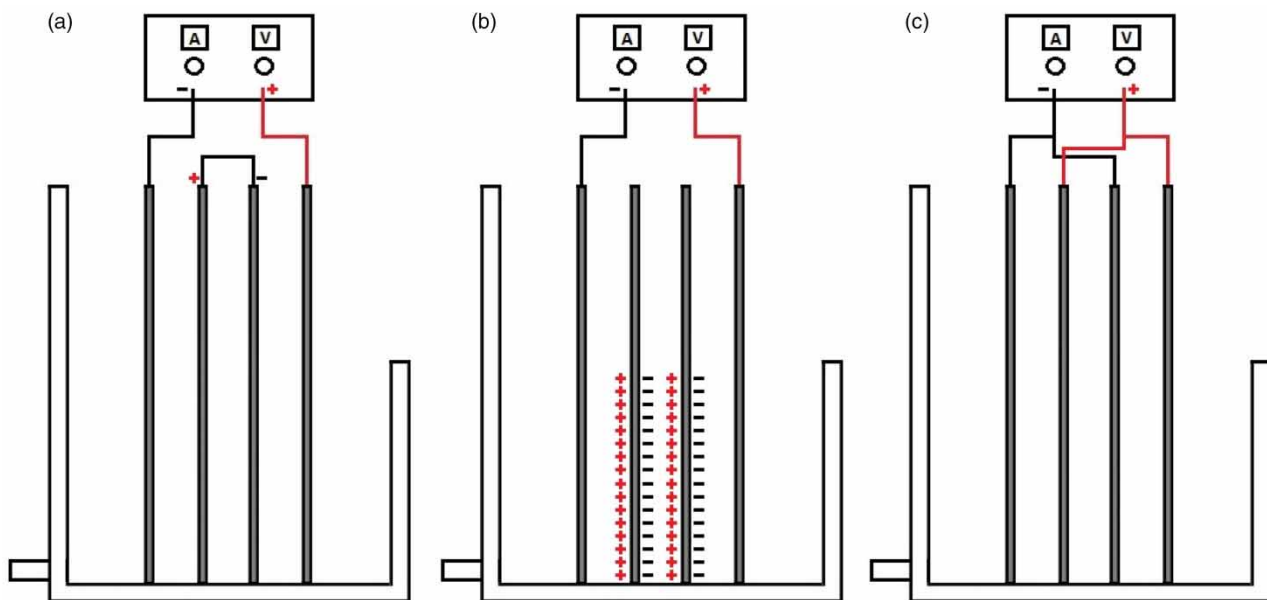


Figure 2 | Electrode's configuration: (a) series monopolar (SM); (b) series bipolar (SB); (c) parallel (P).

removal obtained from the cumulative concentration of fluoride at the outlet of the unit:

$$Removal (\%) = 100 \times \left(1 - \frac{\int_0^t C_{out} dt}{tC_F} \right) \quad (1)$$

where C_F and C_{out} are the concentrations of fluoride in the water entering and leaving the unit, respectively.

In the present work, instead of considering only the final performance of the process in each of the 45 experiments, the values of outlet concentration and voltage in each of the six sampling times were considered, resulting in a dataset containing 270 experimental points. The implementation of the ANN using the obtained dataset considered the input variables: applied current intensity (I), flow rate (Q), feed concentration (C_F), operation time (t), and electrodes configuration (SM, P, SB). The output variables considered were the fluoride concentration in the unit outlet (C_{out}) and the applied voltage (U). The electrodes configuration is a categorical variable that, to be used as input in an ANN, needs to be converted into a numerical variable. The electrodes configuration conversion into a numerical variable was performed by dividing the variable into three variables (SM, P, SB), attributing the value of 1 when the variable corresponds to the configuration used and 0 otherwise. After this transformation, the dataset will be constituted by seven input variables.

The values of the variables present in the dataset vary across different ranges, which can be difficult for the learning process of the ANN. Normalizing all the input variables to the same range can improve the training process performance and is a recommended practice (Rahmanpanah *et al.* 2020). In the present work, the input variables were normalized using the mean-std method. This method uses the mean and the standard deviation of each variable to perform its normalization using the following equation:

$$x' = \frac{x - \bar{x}}{\sigma_x} \quad (2)$$

where x is the originally recorded variable during the experiments, \bar{x} is the mean value of the recorded variable, and σ_x is its standard deviation.

It has been shown that the accuracy of the model and its training performance can be improved by introducing some level of noise to the trained ANN (Zur *et al.* 2009; Noh 2017). This principle works by reducing the sensitivity of the ANN output to small variations in the input variables. This aspect is relevant since the measurements of the different variables during the experiments are not noise-free. Teaching the network not to change its output in a range around the exact input makes the predictions less sensitive to small fluctuations in the input variables. A common procedure is the introduction of Gaussian noise to the input variables during the ANN training. The introduction of noise is achieved by (Zadpoor *et al.* 2013; Rahmanpanah *et al.* 2020):

$$Input' = Input + N\left(0, \frac{Input}{\lambda}\right) \quad (3)$$

where $N(\mu, \sigma)$ represents a normal distribution with mean $\mu = 0$ and standard deviation $\sigma = Input/\lambda$. The signal-to-noise ratio is represented by λ . In the present work, the signal-to-noise value used was 100.

2.3. Artificial neural network (ANN) implementation

In the present work was used a feedforward ANN, also called a multilayer perceptron (MLP). This kind of ANN model is constituted by an input and an output layer and one or more hidden layers between them. The number of neurons in the input layer is equal to the number of input variables. The number of neurons in the output layer is equal to the number of output variables. The number of hidden layers and the number of neurons in each hidden layer are two of the most influential ANN architecture factors. The selection of the appropriate number of hidden layers and the number of neurons in each layer is crucial because an insufficient number of neurons leads to a model unable to capture the nonlinear relationships in the input data (Ripley 2007). On the other side, too many neurons or hidden layers can lead to memorization of the data, resulting in a loss of the ANN generalization ability (overfitting).

The rectified linear unit (ReLU) activation function (Equation (4)) was used in the ANN model in the present work. ReLU is one of the most used activation functions, providing a good predictive ability of the model. Additionally, since the ReLU

model output is equal to 0 for an input less than zero and otherwise is equal to the input, it creates a natural drop-out of the model during the training (Kessler *et al.* 2017).

$$\varphi(x) = \max(0, x) \quad (4)$$

The network training results from the adjustment of the weights and bias until the network output provides a good approximation of the expected values of the training dataset. The most common training algorithm is backpropagation. During the feedforward step, the network is fed with the input elements, and the resulting outputs are compared with the expected values. This step is followed by the backpropagation step, where the weights and bias are adjusted, minimizing the error between the output of the network and the expected value. The present work used the adaptive moment estimation (ADAM) optimization algorithm to perform the ANN parameters adjustments. ADAM uses the gradient of the cost function, then estimates the first and second moments, and applies a bias correction (Kingma & Ba 2014).

Each cycle of feedforward and backpropagation of the entire dataset is called an epoch. The training is carried out during several epochs until the error between the output and target is minimized; at that point, the weights and bias of the ANN are fixed. The final step is to evaluate the performance with the testing dataset.

Some regularization techniques can be used to avoid the over-fitting of the model. One of them is the drop-out method (Hinton *et al.* 2012). This method acts over each layer by randomly deactivating some neurons, being the number of deactivated neurons dependent on the drop-out rate defined for each layer. Another strategy to prevent over-fitting is the early stop method (Samanta *et al.* 2004). In this method, an initially large number of training epochs is specified, and the training is stopped once the model performance on the validation dataset stops improving. Both methods were used in the present work for the training of the ANN.

2.4. Genetic algorithm implementation

In the present work, the real coded version of the GA was used, where vectors of real numbers represent the variables corresponding to the possible solutions (population):

$$x_i = [x_{i1} \ x_{i2} \ \dots \ x_{in}] \quad (5)$$

where the member of the population x_i is constituted by n variables.

After randomly generating the initial population, the GA algorithm uses three main steps to produce the next generation of individuals from the present population: selection of the parents, crossover, and mutation.

The selection of the parents involves the choice of a set of the best individuals of the population. All the individuals of the population must be quantitatively evaluated. This evaluation is made using the objective function, also referred to as the fitness function. After evaluating the population, a process of selecting the parents to produce the next generation is applied. The present work used the roulette wheel algorithm, where the probability of an individual being selected is proportional to its fitness ratio, which is the ratio of the individual fitness and the value of the fitness of the entire population. The selected individuals will be then subject to a crossover process resulting in new individuals containing a combination of the information of both parents. In the present work, the breeding process was performed using the uniform crossover process, where each two selected parents:

$$x_1 = [x_{11} \ x_{12} \ \dots \ x_{1n}] \quad (6)$$

$$x_2 = [x_{21} \ x_{22} \ \dots \ x_{2n}] \quad (7)$$

are combined to produce two children:

$$y_1 = \alpha_1 x_1 + (1 - \alpha_1) x_2 \quad (8)$$

$$y_2 = \alpha_2 x_2 + (1 - \alpha_2) x_1 \quad (9)$$

where α_i is a random number between 0 and 1, generated for each crossover operation.

Finally, the mutation process occurs, where each variable of the new individuals (y_{ij}) has a probability of a random change:

$$y'_{ij} = y_{ij} - \delta \tag{10}$$

where δ is a random number obtained from the normal distribution $N(0, \sigma^2)$, being the variance σ^2 called mutation step. The introduction of mutation increases the GA's exploratory nature; however, excessive mutations can make the convergence of the algorithm difficult (Marghany 2020).

The resulting population containing the old and the new generations is sorted according to each individual's fitness, and the individuals placed at the bottom of the population are eliminated. This process is repeated for several generations until a convergence criterion is met.

3. RESULTS AND DISCUSSION

3.1. Hyperparameters tuning

In the ANN model training process, its parameters (weights and bias) are adjusted by learning algorithms such as backpropagation. However, other model parameters related to its architecture (i.e., number of hidden layers, epochs, learning rate, etc.), known as hyperparameters, should be determined since these parameters control the learning process. Therefore, the correct tuning of the hyperparameters ensures an optimal and reliable training process. The hyperparameter tuning method Hyperband algorithm (Li et al. 2017) was used to find a suitable architecture using the open-source library Keras-tuner (Chollet 2015). The Hyperband algorithm works by allocating a budget (i.e., number of training epochs) to a randomly selected set of ANN model configurations and discarding the low-performance models by applying successive halving (Jamieson & Talwalkar 2016). At each iteration, the budget is redistributed by the remaining models, and the process is repeated until the configuration with the best performance is found. In the present work, the Hyperband algorithm was used to tune the hyperparameters: learning rate, number of hidden layers, number of neurons in each hidden layer, and the drop-out rate. Table 1 presents the search space used on the Hyperband algorithm and the best model obtained.

3.2. Training ANN

A schematic representation of the steps involved in the implementation and training of the ANN is presented in Figure 3.

The training process requires a set of different input data points and their correspondent outputs. The input matrix is constituted by seven columns vectors, and the output matrix by two:

$$Input = \begin{bmatrix} I_1 & Q_1 & C_{F,1} & t_1 & SM_1 & P_1 & SB_1 \\ I_2 & Q_2 & C_{F,2} & t_2 & SM_2 & P_2 & SB_2 \\ \vdots & & & & \ddots & & \vdots \\ I_n & Q_n & C_{F,n} & t_n & SM_n & P_n & SB_n \end{bmatrix} \tag{11}$$

Table 1 | Hyperparameter tuning search space and best model obtained

Parameter	Search space	Best model
N° hidden layers	1-3	2
Learning rate	10^{-2} , 10^{-3} , 10^{-4}	10^{-2}
<u>Hidden layer 1</u>		
n° neurons	1-100	94
Dropout rate	0-0.5	0
<u>Hidden layer 2</u>		
N° neurons	1-100	85
Dropout rate	0-0.5	0
<u>Hidden layer 3</u>		
N° neurons	1-100	-
Dropout rate	0-0.5	-

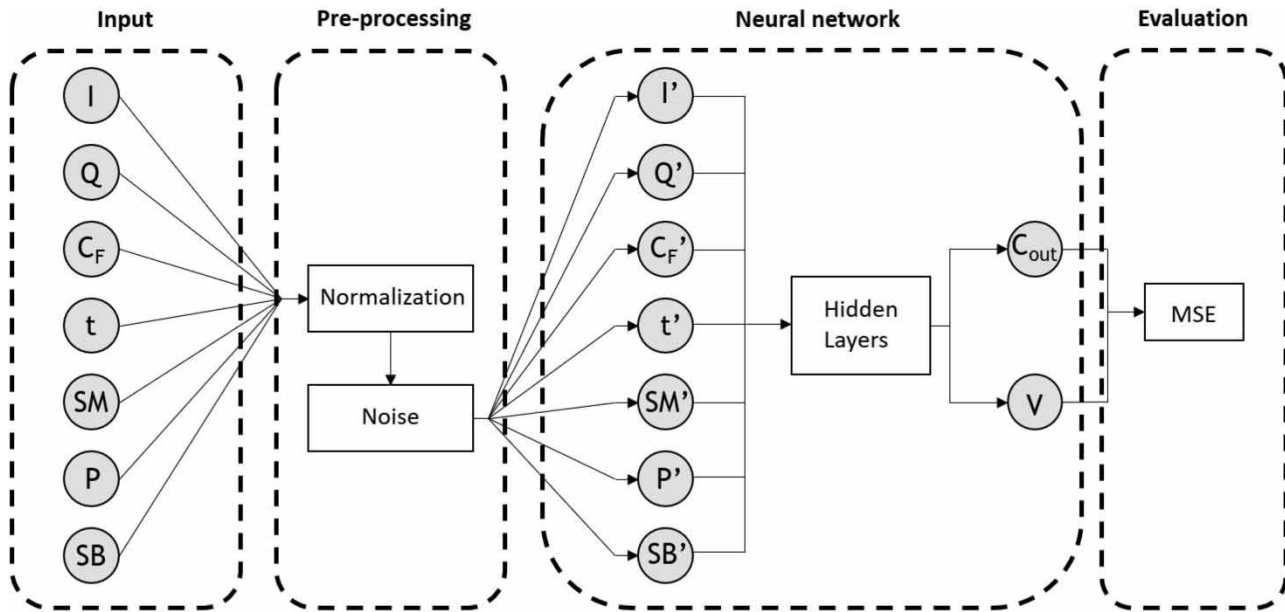


Figure 3 | Schematic representation of the steps involved in the ANN implementation.

$$Output = \begin{bmatrix} C_{out,1} & V_1 \\ C_{out,2} & V_2 \\ \vdots & \vdots \\ C_{out,n} & V_n \end{bmatrix} \quad (12)$$

where n represents the number of data points.

In the pre-processed dataset, 70% of the data points were randomly sampled to be used as the training set, and the rest were used as test and validation sets. The comparison between the output of the model and the training, validation, and testing sets values was made by calculating the mean squared error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (13)$$

where y_i is the measured output and \hat{y}_i the predicted, and n is the number of examples in the dataset.

The performance of the ANN during training can be monitored by the learning curve (Figure 4). The learning curve shows the variation of the MSE of the training and validation sets, calculated at the end of each epoch during the training session. This analysis allows monitoring of the diminishing and convergence of the errors to an acceptable value. After 862 epochs, the model presented an overall MSE of 0.0088 for the training set and 0.0206 for the validation set.

After the training, the model's predictive ability is tested by plotting the target and the predicted values in a $y = x$ plot. Therefore, the closer the scatter points are from the equality line, the more precise the model's predictions. The prediction quality can be quantified by determining the linear regression coefficient for each data set and output; in the results presented in Figure 5, both predictions for voltage and output concentration for the testing set present a determination coefficient (R^2) of 0.999. These results are a crucial aspect since the testing dataset was not involved in the training process. Moreover, the value of R^2 is higher than the value obtained for the same process using statistical models ($R^2 = 0.980$) (Graça *et al.* 2019b), showing the best performance of the ANN model.

Another way to assess the model's performance is by analyzing how the prediction errors are distributed. The histogram of the errors between the target values, and the predicted outputs for both training and test set is presented in Figure 6. This analysis shows that the errors are approximately symmetrically distributed in both datasets around zero in a relatively narrow interval.

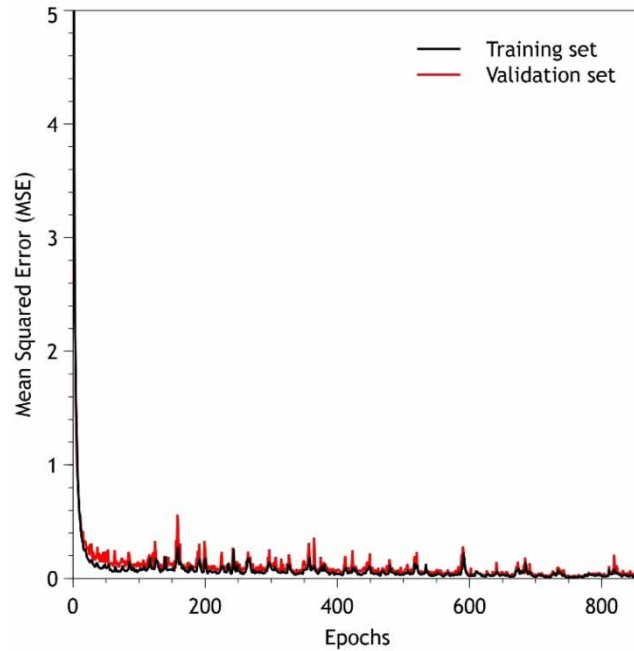


Figure 4 | The learning curve for the ANN training.

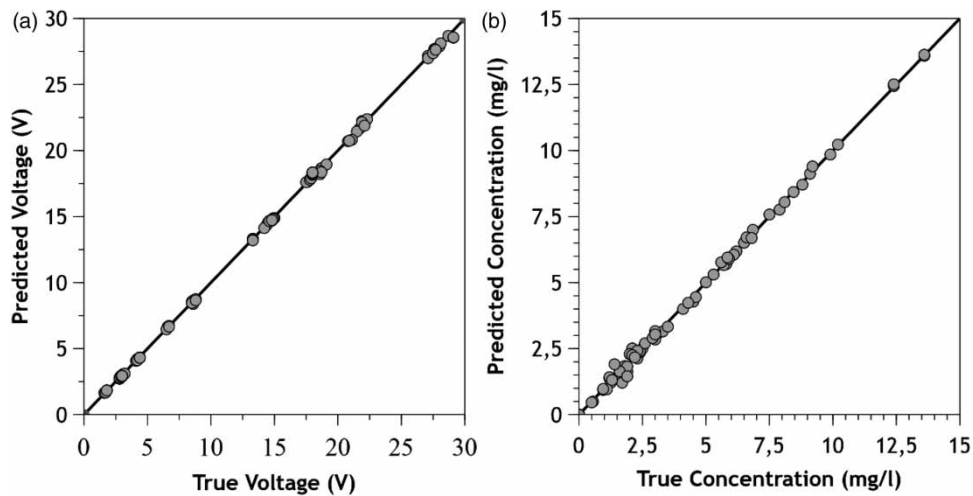


Figure 5 | Scatter plot representation of the predict vs. true values of the testing set for: (a) voltage; (b) outlet concentration.

Figure 7 presents a comparison between the outlet concentration and applied voltage obtained in some CEC experiments, with the values predicted by the trained ANN. The ANN predicted results provide a good fit with the experimental data. Moreover, considering that 30% of the experimental data points were not used during the ANN training, it shows a good predictive ability.

3.3. Optimization of the CEC process

In the present work, the cost associated with the operation was chosen to be the fitness function:

$$Cost = \frac{a \cdot EC + b \cdot m_{Al}}{V_t} \quad (14)$$

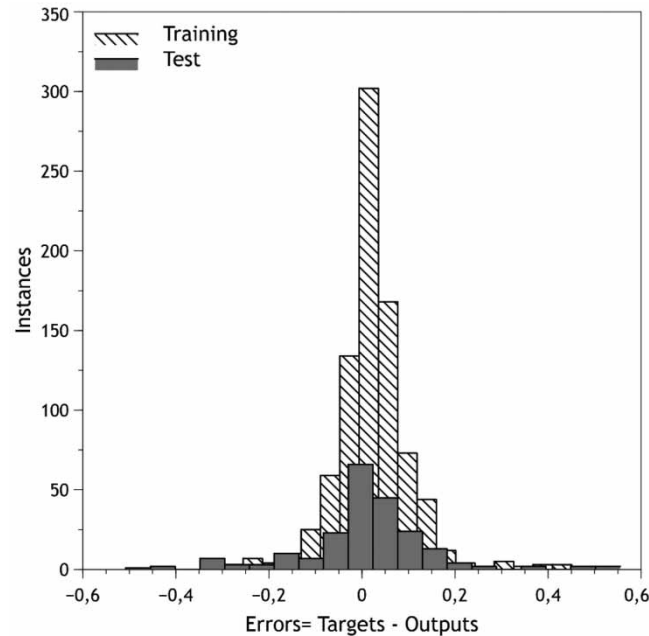


Figure 6 | Histogram of the prediction errors showing the number of occurrences (instances) of an error interval for both the training and the testing sets.

where the cost (€/L) is calculated based on the cost associated with the energy consumption (EC), and the cost of the aluminum dissolved during the electrocoagulation process (m_{Al}). These calculations consider the price of energy ($a = 0.2126$ €/kWh) given by the average cost of energy in the EU and the price of aluminum ($b = 1.659$ €/kg) on the international market in February of 2021. The cost of the continuous electrocoagulation was determined considering a treated volume (V_t) of 5 liters.

The energy consumption was determined using the following expression:

$$EC = I \int_0^t U dt \quad (15)$$

where U is the value of the voltage applied to the electrodes during the CEC operation, I is applied current intensity, and t is the time necessary to treat the volume of the water considered.

The mass of aluminum dissolved from the anode can be determined based on Faraday's law:

$$m_{Al} = \frac{M_{Al}It}{nFz_{Al}} \quad (16)$$

where z_{Al} is the valence number of the aluminum ($z_{Al} = 3$), F is Faraday's constant ($94,485$ C mol⁻¹), M_{Al} is the aluminum molar mass (26.98 g mol⁻¹), t is the operation time, and n is the number of electrodes.

From the optimization point of view, the best fit individuals of the population are the ones with the lowest cost. However, to evaluate the performance of the CEC unit, the removal of fluoride should also be considered. This evaluation was done by calculating the removal (Equation (1)) of each individual and excluding those with a value lower than a pre-set limit. This exclusion process occurs during the generation of the first population and on the individuals of the new generations. Therefore, it is guaranteed that every individual of the population at each generation has a removal above the pre-set limit.

The variables used in the present work for the GA implementation were the current intensity, flow rate, and feed concentration, which are continuous variables limited by upper and lower bounds. The electrode configuration was used as a

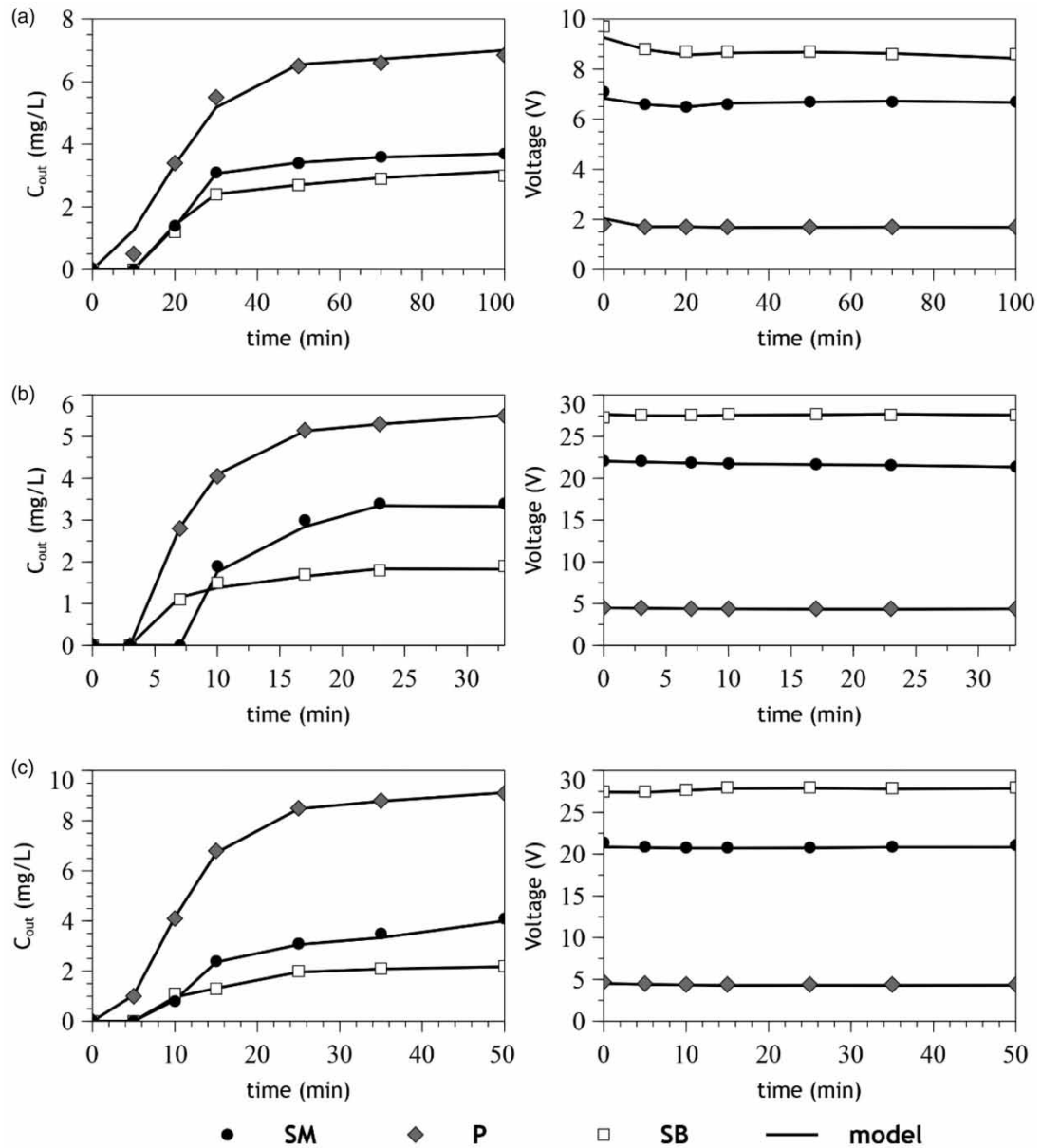


Figure 7 | Comparison between the experimental values of outlet concentration and applied voltage with the values predicted with the ANN for the operating conditions: (a) $I = 40$ mA, $Q = 50$ mL/min, $C_F = 10$ mg/L; (b) $I = 160$ mA, $Q = 150$ mL/min, $C_F = 10$ mg/L; (c) $I = 160$ mA, $Q = 100$ mL/min, $C_F = 15$ mg/L; for the different electrode configurations: series monopolar (SM), series bipolar (SB) and parallel (P).

categorical variable with three possible values. The operating time was used as a dependent variable calculated from the treated volume and the flow rate. The parameters used in the GA implementation are summarized in [Table 2](#).

The optimization of the CEC unit to remove fluoride from water consisted of finding the operating parameters that provided the lowest cost for a pre-set removal. First, the process was optimized considering the minimal removal necessary to obtain treated water with a fluoride concentration below the permissible limit (1.5 mg/l). The removal necessary to achieve this goal will depend on the feed concentration. Second, the process was optimized considering the maximum removal possible. In this calculation, the minimum limit of detection of the analytic method (0.1 mg/l) was considered. The optimization results are presented in [Table 3](#).

The results obtained with the GA show that the optimal conditions that optimized the operating cost are highly dependent on the feed concentration and on the fluoride removal requirement. The interpretation of the results requires an

Table 2 | Parameters of GA

Parameter	Value
N° of variables	5
Initial population	100
Mutation probability	0.2
Mutation step	10
N° of children	2
Lower bonds (I, Q, and C _F)	40, 50, 5
Upper bonds (I, Q, and C _F)	160, 150, 15
Categorical variable ([SM P SB])	[1 0 0], [0 1 0], [0 0 1]
Dependent variable (t)	V _t / Q

Table 3 | Optimized operating conditions obtained by the GA for different feed concentrations and removal requirements

C _F (mg/L)	Maximum removal			Minimal removal		
	5	10	15	5	10	15
I (mA)	90.5	97.0	160.0	75.0	88.3	74.0
Q (ml/min)	150.0	51.2	50.0	150.0	73.6	51.4
Conf. (SM, P, SB)	SM	SB	SB	P	SM	SB
Removal (%)	98.0	99.0	98.5	70.0	85.0	90.0
Cost (€/L)	0.029	0.121	0.316	0.004	0.050	0.074

understanding of the effect of the different operating variables on the CEC unit performance. The process variables current intensity and flowrate have adversarial effects on the optimization of fluoride removal and operating cost. The increase of current intensity will improve the removal due to the increase of the anode's dissolution rate, and consequently, the increase of soluble and insoluble metal hydroxide species available to participate in fluoride retention (Mollah *et al.* 2004). Moreover, the increase of the current intensity enhances the hydrogen bubble production rate and the growth and size of the coagulant flocs, which can improve the removal efficiency (Bouguerra *et al.* 2015). However, high current intensities lead to high energy consumption that, combined with a high anode dissolution rate, increases operating costs. The removal can also be improved by decreasing the flow rate, which increases the residence time inside the unit and enhances the contact time between the fluoride and the generated coagulant. However, the operating time increase is proportional to the increase in energy consumption and, consequently, high operating costs. Therefore, it is necessary for each case to find the right operating conditions that provide an optimized balance between the process removal and its operating cost. It is important to notice that in some cases the values of the optimized variables are the boundaries of its search space, suggesting that better optimal conditions could eventually be found by using large ranges for these variables. However, the use of variable ranges large than the ones used in the training of the ANN leads to extrapolation of the results; therefore, in the present work, the ranges of the variables used were the same used for the ANN training.

The effect of the removal requirement on the CEC optimized operating cost using different electrode configurations is shown in Figure 8. The results show that the operating cost increases with the increase of the fluoride removal requirement, and the electrode configuration that provides the lower cost depends on this parameter. To achieve the pre-set removal requirement while optimizing the operation cost, it is necessary to find the correct balance between the coagulant dosage, which is controlled by the current intensity, and the residence time, which is controlled by the flow rate. The configurations SM and P seem to provide the best result for lower removal requirements, and the SB is the best for higher removals. Considering only the fluoride removal efficiency, the SB configuration presents the best performance for the CEC process (Graça *et al.* 2019b). This superior performance is explained by the contact of the water with alternated polarities when flowing

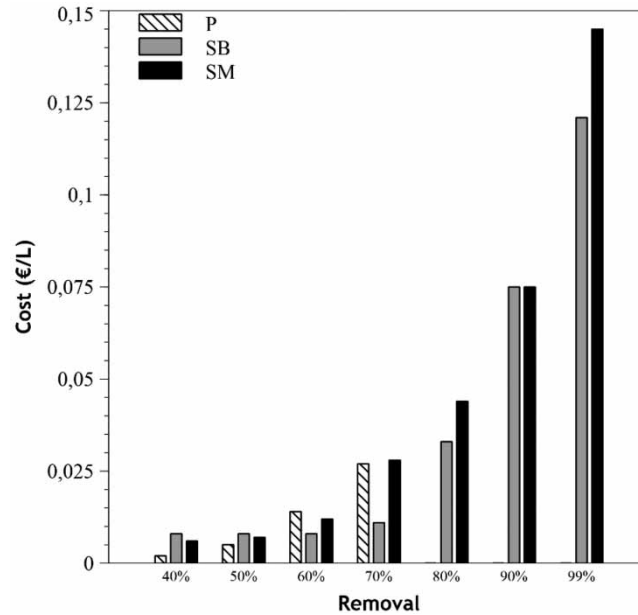


Figure 8 | Effect of the removal requirement on the optimized operation cost for each configuration considering a feed concentration of 10 mg/l.

through the electrodes, which enhances the electro-condensation effect (Ming *et al.* 1987; Hu *et al.* 2003). However, due to the lack of connection between the inner electrodes, the SB configuration requires higher current intensities.

The analysis of the effect of the feed concentration on the optimized operating cost is presented in Figure 9. The operating cost increases due to the higher current and/or lower flow rate necessary to deal with higher fluoride concentrations. Moreover, the increase of fluoride concentration can affect the formation of the aluminum flocs, which can eventually impact the removal performance (Silva *et al.* 2018). These results confirm the suitability of the SB configuration to deal with higher removal requirements and high feed concentrations, suggesting that the SB configuration generally presents a higher removal performance. However, at lower feed concentrations, the high removal performance could not compensate for the high energy consumption associated with the SB configuration, which results from the non-existence of electrical connection between the electrodes.

The choice of the adequate electrode configuration will depend on the kind of application intended for the CEC unit. If the goal is to maximize fluoride removal, the SB configuration is the best choice. However, it is necessary to consider the higher

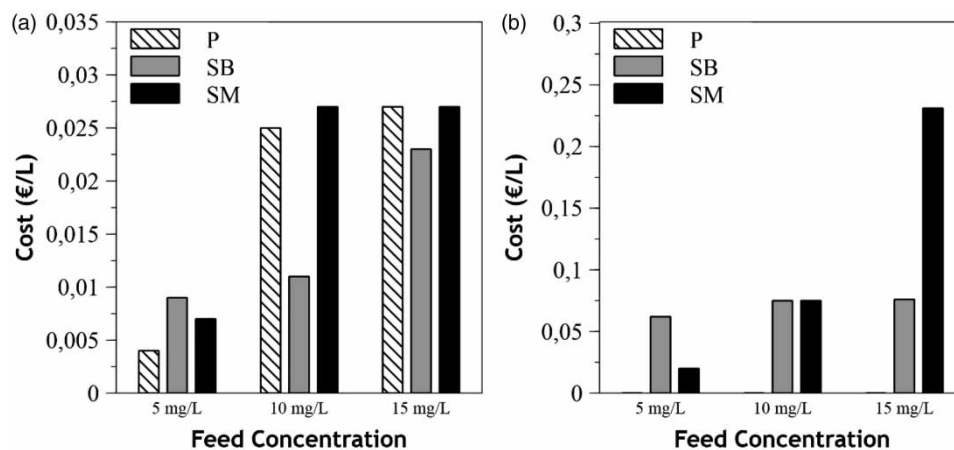


Figure 9 | Effect of the feed concentration of the optimized operation cost for removal requirements of (a) 70%; (b) 90%.

cost associated with this configuration. A possible way to reduce the costs is by reducing the removal requirement or the feed concentration by combining electrocoagulation with other processes, such as adsorption (Barhoumi *et al.* 2019; Khan *et al.* 2019). In this case, the lower cost associated with SM or P configurations could be an essential factor in choosing the best operating conditions.

4. CONCLUSION

In the present work, the CEC process for removing fluoride from the water was modeled using an ANN. The ANN implementation required collecting and pre-processing an experimental dataset containing the operating conditions used as input variables and correspondent outputs related to the unit performance. The choice of the hyperparameters that define the topology of the ANN was performed using the Hyperband algorithm. The trained model presented a good generalization and predictive ability, with an MSE and R^2 on the testing dataset of 0.0206 and 0.999, respectively.

The trained ANN was used to construct the GA algorithm's fitness function. The GA was used to minimize the operating cost of the CEC when limited by a removal requirement. The results showed that the best-operating conditions are highly dependent on the fluoride removal requirement and feed concentration. The results suggest that the best electrode configuration depends on the kind of application intended for the CEC. The operating cost can potentially be substantially reduced by combining CEC with other water treatment processes.

Both ANN and GA were implemented using open-source libraries in the Python programming language, showing great flexibility and efficiency. The availability and no-cost associated with this kind of software make them very attractive tools to be implemented in modeling, simulation, and optimizing environmental and chemical engineering processes.

ACKNOWLEDGEMENTS

This work was financially supported by: Base Funding – UIDB/50020/2020 of the Associate Laboratory LSRE-LCM – funded by national funds through FCT/MCTES (PIDDAC).

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

- Barhoumi, A., Ncib, S., Chibani, A., Brahmi, K., Bouguerra, W. & Elaloui, E. 2019 High-rate humic acid removal from cellulose and paper industry wastewater by combining electrocoagulation process with adsorption onto granular activated carbon. *Industrial Crops and Products* **140**, 111715.
- Barron, A. R. 1993 Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory* **39** (3), 930–945.
- Bock, F. E., Aydin, R. C., Cyron, C. J., Huber, N., Kalidindi, S. R. & Klusemann, B. 2019 A review of the application of machine learning and data mining approaches in continuum materials mechanics. *Frontiers in Materials* **6**, 1–23.
- Borgli, R. J., Stensland, H. K., Riegler, M. A. & Halvorsen, P. 2019 Automatic Hyperparameter Optimization for Transfer Learning on Medical Image Datasets Using Bayesian Optimization. In: *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, pp. 1–6.
- Bouguerra, W., Barhoumi, A., Ibrahim, N., Brahmi, K., Aloui, L. & Hamrouni, B. 2015 Optimization of the electrocoagulation process for the removal of lead from water using aluminium as electrode material. *Desalination and Water Treatment* **56** (10), 2672–2681.
- Cañizares, P., Martínez, F., Rodrigo, M. A., Jiménez, C., Sáez, C. & Lobato, J. 2008 Modelling of wastewater electrocoagulation processes: part I. general description and application to kaolin-polluted wastewaters. *Separation and Purification Technology* **60** (2), 155–161.
- Castañeda, L. F., Coreño, O., Nava, J. L. & Carreño, G. 2020 Removal of fluoride and hydrated silica from underground water by electrocoagulation in a flow channel reactor. *Chemosphere* **244**, 1–7.
- Changmai, M., Pasawan, M. & Purkait, M. K. 2018 A hybrid method for the removal of fluoride from drinking water: parametric study and cost estimation. *Separation and Purification Technology* **206**, 140–148.
- Chollet, F. 2015 Keras: Deep Learning Library for Theano and TensorFlow. Available from: <https://keras.io/> (accessed on 23 February 2021).
- Deepthi, B. P., Shreyas, B. V. & Vishwanath, K. N. 2021 Defluoridation of Groundwater Using Electrocoagulation Followed by Adsorption. In: *Trends in Civil Engineering and Challenges for Sustainability* (Narasimhan, M. C., George, V., Udayakumar, G. & Kumar, A. eds.). Springer Singapore, Singapore, pp. 525–542.
- Drinking Water Directive 1998 On the quality of water intended for human consumption. *Official Journal of the European Communities* **330**, 32–54.

- Duarte, F. & Ratti, C. 2018 The impact of autonomous vehicles on cities: a review. *Journal of Urban Technology* **25** (4), 3–18.
- Fajardo, A. S., Martins, R. C., Silva, D. R., Martínez-Huitle, C. A. & Quinta-Ferreira, R. M. 2017 Dye wastewaters treatment using batch and recirculation flow electrocoagulation systems. *Journal of Electroanalytical Chemistry* **801**, 30–37.
- Fakhri, A. & Adami, S. 2013 Response surface methodology for adsorption of fluoride ion using nanoparticle of zero valent iron from aqueous solution. *Journal of Chemical Engineering & Process Technology* **4**, 1–6.
- Graça, N. S., Ribeiro, A. M. & Rodrigues, A. E. 2019a Modeling the electrocoagulation process for the treatment of contaminated water. *Chemical Engineering Science* **197**, 379–385.
- Graça, N. S., Ribeiro, A. M. & Rodrigues, A. E. 2019b Removal of fluoride from water by a continuous electrocoagulation process. *Industrial and Engineering Chemistry Research* **58** (13), 5314–5321.
- Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. R. 2012 Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors.
- Hu, C. Y., Lo, S. L. & Kuan, W. H. 2003 Effects of co-existing anions on fluoride removal in electrocoagulation (EC) process using aluminum electrodes. *Water Research* **37** (18), 4513–4523.
- Inguva, P., Bhute, V. J., Cheng, T. N. H. & Walker, P. J. 2021 Introducing students to research codes: a short course on solving partial differential equations in Python. *Education for Chemical Engineers* **36**, 1–11.
- Jamieson, K. & Talwalkar, A. 2016 Non-stochastic best arm identification and hyperparameter optimization. In: *Proceedings of Artificial Intelligence and Statistics*, 9-11 May 2016, Cadiz, Spain. PMLR, pp. 240–248.
- Jones, L. D., Golan, D., Hanna, S. A. & Ramachandran, M. 2018 Artificial intelligence, machine learning and the evolution of healthcare. *Bone & Joint Research* **7** (3), 223–225.
- Kessler, T., Dorian, G. & Mack, J. H. 2017 Application of a Rectified Linear Unit (ReLU) Based Artificial Neural Network to Cetane Number Predictions. In: *ASME 2017 Internal Combustion Engine Division Fall Technical Conference*.
- Khan, S. U., Islam, D. T., Farooqi, I. H., Ayub, S. & Basheer, F. 2019 Hexavalent chromium removal in an electrocoagulation column reactor: process optimization using CCD, adsorption kinetics and pH modulated sludge formation. *Process Safety and Environmental Protection* **122**, 118–130.
- Khan, S. U., Asif, M., Alam, F., Khan, N. A. & Farooqi, I. H. 2020 Optimizing Fluoride Removal and Energy Consumption in a Batch Reactor Using Electrocoagulation: A Smart Treatment Technology. In: *Smart Cities – Opportunities and Challenges* (Ahmed, S., Abbas, S. M. & Zia, H. eds.). Springer Singapore, Singapore, pp. 767–778.
- Khemis, M., Leclerc, J.-P., Tanguy, G., Valentin, G. & Lapique, F. 2006 Treatment of industrial liquid wastes by electrocoagulation: experimental investigations and an overall interpretation model. *Chemical Engineering Science* **61** (11), 3602–3609.
- Kingma, D. P. & Ba, J. 2014 Adam: A method for stochastic optimization. *CoRR*.
- Lacasa, E., Cañizares, P., Sáez, C., Martínez, F. & Rodrigo, M. A. 2013 Modelling and cost evaluation of electro-coagulation processes for the removal of anions from water. *Separation and Purification Technology* **107**, 219–227.
- Lacson, C. F. Z., Lu, M.-C. & Huang, Y.-H. 2021 Fluoride-containing water: a global perspective and a pursuit to sustainable water defluoridation management – an overview. *Journal of Cleaner Production* **280**, 124236.
- Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A. & Talwalkar, A. 2017 Hyperband: a novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research* **18** (1), 6765–6816.
- Lin, C.-J. 2004 A GA-based neural fuzzy system for temperature control. *Fuzzy Sets and Systems* **143** (2), 311–333.
- Marghany, M. 2020 Chapter 10 - Principles of genetic algorithm. In: *Synthetic Aperture Radar Imaging Mechanism for Oil Spills* (Marghany, M. ed.). Gulf Professional Publishing, Cambridge, MA, pp. 169–185.
- Matteson, M. J., Dobson, R. L., Glenn, R. W., Kukunoor, N. S., Waits, W. H. & Clayfield, E. J. 1995 Electrocoagulation and separation of aqueous suspensions of ultrafine particles. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **104** (1), 101–109.
- Ming, L., Yi, S. R., Hua, Z. J., Yuan, B., Lei, W., Ping, L. & Fuwa, K. C. 1987 Elimination of excess fluoride in potable water with coacervation by electrolysis using an aluminum anode. *Fluoride* **20** (2), 54–63.
- Mollah, M. Y. A., Morkovsky, P., Gomes, J. A. G., Kesmez, M., Parga, J. & Cocke, D. L. 2004 Fundamentals, present and future perspectives of electrocoagulation. *Journal of Hazardous Materials* **114** (1–3), 199–210.
- Morales-Rivera, J., Sulbarán-Rangel, B., Gurubel-Tun, K. J., del Real-Olvera, J. & Zúñiga-Grajeda, V. 2020 Modeling and optimization of COD removal from cold meat industry wastewater by electrocoagulation using computational techniques. *Processes* **8** (9), 1–13.
- Nasr, M., Ateia, M. & Hassan, K. 2016 Artificial intelligence for greywater treatment using electrocoagulation process. *Separation Science and Technology* **51** (1), 96–105.
- Nogueira, I. B. R., Ribeiro, A. M., Requião, R., Pontes, K. V., Koivisto, H., Rodrigues, A. E. & Loureiro, J. M. 2018 A quasi-virtual online analyser based on an artificial neural networks and offline measurements to predict purities of raffinate/extract in simulated moving bed processes. *Applied Soft Computing Journal* **67**, 29–47.
- Noh, H. 2017 Regularizing deep neural networks by noise: its interpretation and optimization. *Proceedings Advances in Neural Information Processing Systems* (NIPS 2017), Long Beach, CA, USA, 5109–5118.
- Oliveira, L. M. C., Koivisto, H., Iwakiri, I. G. I., Loureiro, J. M., Ribeiro, A. M. & Nogueira, I. B. R. 2020 Modelling of a pressure swing adsorption unit by deep learning and artificial intelligence tools. *Chemical Engineering Science* **224**, 115801.

- Picos-Benítez, A. R., López-Hincapié, J. D., Chávez-Ramírez, A. U. & Rodríguez-García, A. 2017 Artificial intelligence based model for optimization of COD removal efficiency of an up-flow anaerobic sludge blanket reactor in the saline wastewater treatment. *Water Science and Technology* **75** (6), 1351–1361.
- Podgorski, J. E., Labhasetwar, P., Saha, D. & Berg, M. 2018 Prediction modeling and mapping of groundwater fluoride contamination throughout India. *Environmental Science & Technology* **52** (17), 9889–9898.
- Prakasham, R. S., Sathish, T. & Brahmaiah, P. 2011 Imperative role of neural networks coupled genetic algorithm on optimization of biohydrogen yield. *International Journal of Hydrogen Energy* **36** (7), 4332–4339.
- Rahmanpanah, H., Mouloudi, S., Burvill, C., Gohari, S. & Davies, H. M. S. 2020 Prediction of load-displacement curve in a complex structure using artificial neural networks: a study on a long bone. *International Journal of Engineering Science* **154**, 103319.
- Ripley, B. D. 2007 *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge, UK.
- Samanta, B., Bandopadhyay, S., Ganguli, R. & Dutta, S. 2004 Sparse data division using data segmentation and Kohonen network for neural network and geostatistical ore grade modeling in Nome offshore placer deposit. *Natural Resources Research* **13** (3), 189–200.
- Shen, G. & Liu, Q. 2020 *Performance Analysis of Linear Regression Based on Python*. Cognitive Cities, Singapore, 2020; Shen, J.; Chang, Y.-C.; Su, Y.-S.; Ogata, H., Eds. Springer Singapore: Singapore, 2020; pp. 695–702.
- Silva, J. F. A., Graça, N. S., Ribeiro, A. M. & Rodrigues, A. E. 2018 Electrocoagulation process for the removal of co-existent fluoride, arsenic and iron from contaminated drinking water. *Separation and Purification Technology* **197**, 237–243.
- Sivanandam, S. N. & Deepa, S. N. 2006 *Introduction to Neural Networks Using Matlab 6.0*. Tata McGraw-Hill, New Delhi, India.
- Voulodimos, A., Doulamis, N., Doulamis, A. & Protopapadakis, E. 2018 Deep learning for computer vision: a brief review. *Computational Intelligence and Neuroscience* **2018**, 7068349.
- Wang, C., Ye, Z., Yu, Y. & Gong, W. 2018 Estimation of bus emission models for different fuel types of buses under real conditions. *Science of The Total Environment* **640–641**, 965–972.
- Whitley, D. 1994 A genetic algorithm tutorial. *Statistics and Computing* **4** (2), 65–85.
- World Health Organization 1993 *Guidelines for Drinking-Water Quality*. World Health Organization, Geneva, Switzerland.
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G. S., Hughes, M. & Dean, J. 2016 Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *ArXiv* abs/1609.08144.
- Zadpoor, A. A., Campoli, G. & Weinans, H. 2013 Neural network prediction of load from the morphology of trabecular bone. *Applied Mathematical Modelling* **37** (7), 5260–5276.
- Zur, R. M., Jiang, Y., Pesce, L. L. & Drukker, K. 2009 Noise injection for training artificial neural networks: a comparison with weight decay and early stopping. *Medical Physics* **36** (10), 4810–4818.

First received 16 April 2021; accepted in revised form 23 July 2021. Available online 5 August 2021